Corporate Data Quality Management

From Theory to Practice

Ana Lucas
Laboratório Nacional de Engenharia Civil e
Instituto Superior de Economia e Gestão
Lisboa, Portugal
ana.lucas@lnec.pt

Abstract— **It is now assumed that poor quality data is costing large amounts of money to corporations all over the world. Although research on methods and techniques for data quality assessment and improvement have begun in the early nineties of the past century and being currently abundant and innovative, it is noted that the academic and professional communities virtually have no dialogue, which turns out to be harmful to both of them. The challenge of promoting the relevance in information systems research, without compromising the necessary rigor, is still present in the various disciplines of information systems scientific area [1,2], including the data quality one.**

In this paper we present "data as a corporate asset" as a business philosophy, and a framework for the concepts related to that philosophy, derived from the academic and professional literature. According to this framework, we present, analyze and discuss a single explanatory case study, developed in a fixed and mobile telecommunications company, operating in one of the European Union Countries. The results show that, in the absence of data stewardship roles, data quality problems become more of an "IT problem" than typically is considered in the literature, owing to Requirements Analysis Teams of the IS Development Units, to become a "quality negotiator" between the various stakeholders. Other findings are their bottom-up approach to data quality management, their biggest focus on motivating employees through innovative forms of communication, which appears to be a critical success factor[1] (CSF) for data quality management, as well as the importance of a data quality champion [3] leadership.

Keywords-data quality management; framework; data quality initiative; case study

I. INTRODUCTION

Data quality management (DQM) is an issue of growing importance to the academic and professional communities. Today, there is a great concern about the quality of corporate data, as data of poor quality means inaccurate information, which in turn means a wasting of resources and harming the organization externally, namely through the relationships with its customers.

We define *data* as "stored representations of objects and events that have meaning and importance in the users' environment" and *information* as "data that have been processed in such a way as to increase the knowledge of the person who uses the data [37]. *Knowledge* is "information that changes something or somebody, either by becoming grounds for actions, or by making an individual (or an institution) capable of different or more effective *action* [38]. Knowledge resides in the user and not in the information, and means the user capacity to act using the available information. Although data and information mean slightly different things, for simplicity reasons, and in line with other research approaches to data quality, we will use, in the context of this paper, *data* and *information* interchangeably.

One of the most widespread definitions of data quality states that *data are of high quality* "if they are fit for their intended uses in operations, decision making and planning" [17, 18].

According to [4], the costs of poor data quality can be classified into three categories:

- *Process failure costs*, such as costs associated with misdelivered or nondeliverable mail due to inaccurate mailing addresses;
- *Information scrap and rework*, such as costs associated with resending correspondence;
- *Opportunity costs*, due to the lost and missed revenues. For example, due to low accuracy of costumers' addresses associated with "loyalty cards", a percentage of those cards' owners are not reached in advertising campaigns, resulting in lower revenues.

Just to get an idea of poor data quality costs, it was estimated in [6] that current data quality problems cost U.S. businesses more than USD 600 billion a year. The report's findings were based on interviews with industry experts, leading-edge customers, and survey data from 647 respondents.

DQM can be defined as quality-oriented management of data as an asset [7], that is the "the application of total quality management (TQM) concepts and practices to improve data

[1]According to [5] "The critical success factors are the characteristics, conditions, or variables that when properly sustained, maintained or managed, can have a significant impact on the success of an organization, competing in a particular industry."

and information quality, including setting data quality policies and guidelines, data quality measurement (including data quality auditing and certification), data quality analysis, data cleansing and correction, data quality process improvement, and data quality education"[40]. To be effective, data quality management must go beyond the activities of *fixing non-quality data*, to *preventing data quality problems* through managing data over its lifecycle to meet the information needs of data stakeholders. Moreover, DQM requires *breaking down the stovepipes separating data across business units and creating collaboration between business and IT functions*, in order to address both organizational and technical perspectives, requiring thus a *profound cultural change demanding leadership, authority, control and allocation of resources,* which means governance, specifically *data governance*. With data governance, companies are able to implement corporate-wide accountabilities for DQM, encompassing professionals from both business and IT departments. Although there is collaboration between business and IT departments up to a certain extent, IT is often left in the lurch when it comes to improving data quality (DQ) and managing corporate data [9] cited by [7].

Our research questions being *why do corporations engage themselves in a data quality management initiative* or, put another way, which are the drivers to such initiative, *and how do they implement DQM.*

Being at the very beginning of this investigation, we decided that our research strategy should begin by working on a single explanatory case study [10], the goal being to understand, in its context, *why* a specific company decided to embark in a data quality management initiative, *how* it did that and *which objectives* have been achieved, so far, as well as comparing what we have found in the case to theories from the literature.

This paper is organized as follows: in this section the research problem is introduced, in the background section we present a framework for data management related concepts and functions, the data quality concept and some of its most important dimensions, as well as some approaches and methodologies used by organizations to assess and improve the quality of their data, followed by the case study presentation. In the analysis and discussion section we analyze and discuss the findings and finally, we present the limitations of the work and some guidelines for future research.

II. BACKGROUND

This section involves a presentation of the fundamentals underlying the work, namely the presentation of a framework for the concepts related to the data management, the data quality concept and some of its most important dimensions, as well as some approaches and methodologies used by organizations to assess and improve the quality of their data.

A. *Data Management Approach*

The *data management approach* comprises all the disciplines and functions related to managing data as a valuable resource. According to [40], *data management* is:

1. The business function that develops and executes plans, policies, practices and projects that acquire, control, protect, deliver and enhance the value of data and information;
2. A program for implementation and performance evaluation of the data management function;
3. The field of disciplines required to perform the data management function;
4. The profession of individuals who perform data management disciplines;
5. In some cases, a synonym for a Data Management Services Organization that performs data management activities.

Literature, whether from academic or professional sources, presents a set of concepts, roles and responsibilities along with their definitions, related to the management of data considered as a corporate asset. In order to clarify and organize these concepts, we define a philosophy underlying all of them, and present, in Table I, a (not exhaustive) list of concepts, roles and responsibilities related to the data management approach and in Fig.1 their main interrelationships.

TABLE I. CONCEPTS AND FUNCTIONS FOR THE DATA MANAGEMENT APPROACH

Concept/Role or Responsibiliy	Definition
Corporate Data Philosophy (CDPh)	*Corporate Business Philosophy* is a long term corporate vision and consists of a set of values that have to be considered above all kinds of policies, strategies, roles or technologies. A *Corporate Data Philosophy* (CDPh), in line with Corporate Business Philosophy, considers data as a corporate asset across the organization, which means, it turns its focus away from the expense associated with acquiring, managing and storing data, towards the business value that can be obtained from using the data and its full strategic lifecycle [12]
Corporate Data Policy (CDP)	CDP recognizes CDPh, and defines the broad guidelines governing data, such as: a) data must be shared and reused in order to support cross-process integration or, put another way, transforms data property from being departmental or personal to being corporate property; b) prescribes the maximization of the value created by data assets
Corporate Data Strategy (CDS)	A CDS is a long term plan of action designed to achieve the directions prescribed by CDP in line with *Corporate Business Strategy*
Data Governance (DG)	– DG is the exercise of authority, control and shared decision-making (planning, monitoring and enforcement) over the management of data assets. Data Governance is high-level planning and control over data management and coordinates the collaboration between IT and the business[13] – In our opinion, DG mainly refers to: a) *strong leadership over the management of data assets*; b) *defining corporate data strategy* (CDS), in line with CDP; c) *providing resources and organizational structures* that enable the

Concept/Role or Responsibiliy	Definition
	implementation of the strategy and goals d) *cascading CDS and goals down into the organization*
Data Quality Management (DQM)	– DQM consists of the planning, implementation and control activities, which are usually supported by a *data quality methodology (DQm)* and applies appropriate *quality management techniques (DQT) and tools (DQt)* to measure, assess, improve and ensure the quality of data. *(*Adapted from [13]). – Moreover, adapting to *data quality* the vocabulary presented in [39], we can define DQM as as a set of coordinated activities to direct and control an organization with regard to data quality
Data Quality Methodology (DQm)	A DQm is a set of guidelines and techniques that, starting from input information describing a given application context, defines a rational process to assess and improve the quality of data [14]. A DQm is made of phases and activities and uses *techniques* (DQT) and *tools* (DQt) to accomplish its work
Data Quality Techniques (DQT)	DQTs can be *data* and *process driven*. The *data driven* DQTs correspond to algorithms, heuristics, knowledge-based procedures and learning processes that provide a solution for specific DQ problems [15], like *record linkage* (eg finding and merging duplicates, ie, different records that represent the same real world entity) or *standardization techniques* (comparing data with lookup tables, and updating it accordingly). *Process driven* techniques are used to describe, analyze and reengineer the information production processes, and they are mainly of two types: *process control* and *process redesign* [15]
Data Quality Tools (DQt)	DQt are software products that implement specific DQTs, particularly data driven techniques, such as *profiling, parsing and standardizing, generalized "cleansing", matching, monitoring* and *enrichment.* (adapted from [16])
Data Steward (DS)	A DS is a data caretaker for some *line of business* (LOB), who must ensure that the interests of those LOBs are reflected in the data content and quality rules definitions, despite addressing solutions that must work across the organization
Data Quality Champion (DQC)	According to [3], DQC are managers who actively and vigorously promote their personal vision for using data quality related technology innovations. They push projects over approval, provide political support, keep participants informed, and allocate resources to data quality projects

Fig.1 presents the main data management related activities at the strategic (first two columns), tactical and operational levels, as well as the corresponding *enablers* and *roles and responsibilities* involved in each level.

Organizations that consider data as a corporate asset and have a quality culture, tend to run activities in all the three levels, while those who are just solving data quality problems in specific datasets, tend to run only activities at the operational and perhaps tactical level, which is what is happening, as we shall see, in our case study.

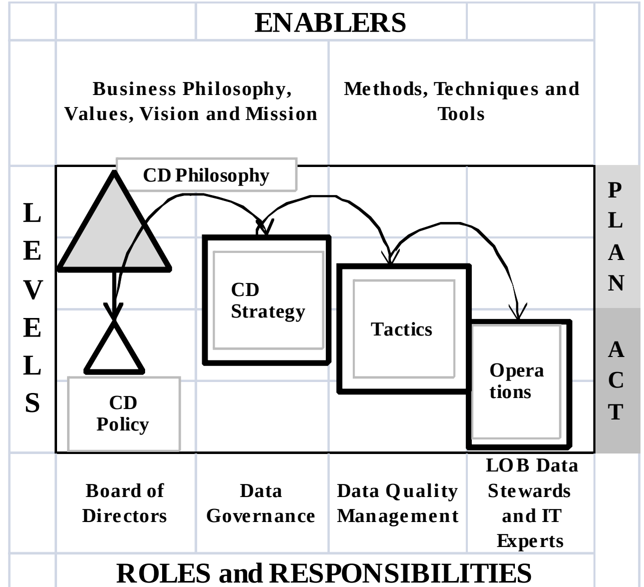

Figure 1. *Data Management Approach* Framework

B. *The Data Quality Concept*

Data are of high quality "if they are fit for their intended uses in operations, decision making and planning" [17, 18].

According to [19] data result from interplay of two components: data models and data values. The data model is simply a structured view of the real world that recognizes entities such as objects or ideas in the real world, attributes such as properties of entities and associations between entities. The data values are the realization of a specific attribute of an entity or a particular association.

Such as the quality of a product depends on the process by which the product is designed and produced, the quality of data depends on the design and production processes involved in their generation. To design for better quality, it is foremost necessary to understand what quality means and how it is measured. Data quality, as presented in the literature, is a multidimensional concept [20, 21].

The *data quality dimensions* are the characteristics of data quality that are meaningful for data consumers, namely those related to data models and data values. Although we intuitively associate data quality with the data values' intrinsic characteristics, such as accuracy, we easily conclude, by using the above concept, that there are other meaningful characteristics of data quality that are prized by their users, such as timeliness, relevancy, etc. With regard to data models' dimensions, the research literature mainly considers intra-data model's dimensions (those related to a specific database schema), although some work is already dealing with the inter-data models issues [26] or, put otherwise, issues that apply to *corporate data architecture*. Those research issues will support, theoretically, the assessment and improvement of corporate data integration, and the much publicized practitioners' concept of *Master Data Management* (MDM)[2].

Although there is no consensus among the various proposals, neither on the number of dimensions, nor on their definitions and metrics (when available), there are four dimensions, perhaps the most significant ones, although with small differences in definition, that are common to the three main proposals [21, 22, 23]: *accuracy*, *completeness*, *timeliness* and *consistency*. These four dimensions should, according to our experience and research [41], be complemented with the *relevancy* dimension, which is proposed by the empirical approach [21].

Moreover, we are convinced, based on our own research, as well as on other researchers' work, that data quality dimensions and their relative importance are highly dependent on the specific field of application.

References [14, 15] provide comparable definitions for data quality dimensions, and they present and discuss other dimensions' definitions available from the literature.

C. Data Quality Management Maturity Models

An organization must move from seeing data quality as an initiative that provides few benefits to something that raises the value of a core asset to the organization, according to CD philosophy. As attitudes change, so does the value proposition for data quality, as it moves from being viewed as a cost that could be eliminated to be considered a strategic initiative within the corporation.

There has only been limited research on instruments to assess the progress and performance of DQM initiatives, usually named *data quality management maturity models*[3], the exception being, to our best knowledge, the models developed in [24], [25] and [26]. Furthermore, there are some well publicized practitioners' approaches, like the ones from [27, 28, 42], although they are lacking an underlying theory base. In the context of this work, we will use two important stages coming from the consultants' side[4] [28], because they are widely used by practitioners, namely in our case study context. These stages are called *reactive* and *proactive*, although they are not disjoint, as one organization can be simultaneously in both stages, despite the objective being to attain the proactive and the governed ones.

Roughly speaking, an organization is at the *reactive stage* when it is making efforts to fix data problems and typically resorts to tools (DQt) to do so and it is at the *proactive stage* when it considers the data as a strategic asset changing from a posture of fixing problems to preventing them [28]. To achieve this stage it is necessary executive sponsorship as well as IT and business staff committed to working together [28]. As will be noted below, MyTelecom is simultaneously in the reactive and proactive stages, as it is, at the same time, fixing problems on previously existing data and modifying business processes so that the new data may have better quality.

D. Methodologies for Data Quality Assessment and Improvement

According to [14] *data quality methodologies* (DQm) apply two types of *strategies* in their improvement activities: *data-driven* and *process-driven*, although some of them adopt *mixed* ones. Roughly speaking, *data-driven* strategies improve the quality of data by directly modifying their value, and *process-driven* strategies improve quality by redesigning the processes that create or modify data.

Although the various methodologies use different strategies, phases, activities and data quality dimensions, they ordinarily have two main common phases: *assessment* and *improvement*. In the *assessment phase*, a diagnosis of data quality, along with relevant quality dimensions, is performed using adequate data quality tools (DQt). *Improvement* mainly concerns: a) the identification of the causes of errors; b) error correction using appropriate DQt and c) redesign, using specific data quality techniques (DQT), of the processes that create or modify data in order to improve their quality. Reference [14] presents and compares some of the most widespread methodologies.

Adapting to *data* the quality vocabulary presented in ISO 9000:2005(E) [39], we consider *data quality assurance* as the part of data quality management[5] focused on providing confidence that data quality requirements will be fulfilled, and *data quality control* as part of data quality management focused on fulfilling quality requirements.

Table II contextualizes some of the above presented concepts.

TABLE II. DQ MANAGEMENT IN CONTEXT

HOW? / WHAT?	**Data Quality Control**	**Data Quality Assurance**
DQM Activities	Fix non-qualty data	Preventing data quality problems from arising
Maturity Stages	Reactive	Proactive
Data Quality Methodologies Strategies	Data-driven	Process-driven

III. THE CASE STUDY

This company is in the fixed and mobile telecommunications business and, for confidentiality reasons,

[2]According to [34], "Master Data Management is a collection of best data management practices that orchestrate key stakeholders, participants, and business clients in incorporating the business applications, information management methods, and data management tools to implement the policies, services and infrastructure to support the capture, integration, and subsequent shared use of accurate, timely, consistent, and complete master data".

[3]In line with the Capability Maturity Model (CMM) developed by the Software Engineering Institute

[4]In general, practitioners usually call them *data governance maturity models*. Although data quality management does not equal data governance, as seen before, the corresponding maturity levels are always strong related

[5]

is being designated by the fictitious name MyTelecom. It operates in one of the European Union Countries and had, in 2010, a turnover of around Eur. 750 million. Its product and services catalog consists of mobile communications services (mobile and Internet) pre and post-paid and fixed (telephone, digital television and Internet), which are offered on optical fiber structure or ADSL (Asymmetric Digital Subscriber Line). The company belongs to a national business group, with interests in multiple sectors all around the world.

Its mission stands out:

- "... Whose ambition is to be the best communications services provider in this country ..."
- "... Striving to consistently create products, services and innovative solutions that fully meet the needs of its markets and generate superior economic value."

The data collection was done through two interviews to the Quality Management Systems (QMS) Coordinator and the observation of data governance reports (whose content is analyzed below) followed some email exchange to clarify some aspects. The QMS belongs to the Quality Technical and Business Information Systems Unit (QTEC&BIS) of the IT Department, and consists of about ten people.

Our respondent is clearly a *data quality champion*, ie, someone that provides political support, keeps participants informed and allocates resources to data quality projects [3]. According to the last reference "the presence of one or more data quality champions is positively associated with management commitment to data quality", which seems to happen in this case.

A. Why (a Data Quality Initiative)?

In November 2007 they started, within the IT[6], the Enterprise Data Governance Initiative, whose driver was the business process of billing and which began to focus on the name, address and zip code customers' attributes. Although there are also many data quality problems at the Data Warehouse (DW), the IT Unit believes that, improving the data at the operational level will resolve, in part, the problems at the DW level.

The company has incurred in extra post office costs due to poor quality of its customer data because, despite having a contract with the post office company stating that they must use all means to deliver their bills, including manual correction, this contract has non-negligible costs. They hope to reduce these costs by improving the quality of customer data as a result of this initiative.

B. How (are they Working)?

As stated before, IT Department is responsible for the data quality management initiative, the main project sponsor being the CIO, although they are "winning" business sponsors at the director's and manager's levels.

[6]The IT unit has about 170 company employees, which must be added by 130 outsourcers.

They do not have, until now, a *Corporate Data Policy* (CDP) nor a *Master Data Management* (MDM) initiative, so they can still have, for instance, the same attribute with different codifications, such as the agent identifier: not being clearly defined the "range" of codes that each department of marketing & sales (Corporate Marketing and SMEs) should use to identify "their" agents, which is a risk and requires manual control that was dispensable. They find it very difficult to implement policies and architectural options, such as CDPs and MDM, in a very competitive industry, where projects have very short time to market.

Asked about the processes used under this initiative our interlocutor identified them on the following order: Awareness, Exploration, Reporting, Fixing and Preventing.

He believes that the initiative was, at first, reactive with the aim to correct customers' attributes. Nevertheless they began analyzing and correcting the root causes of the errors and so they are now, in his words, betting on proactive and reactive processes.

IT people believe that DQ problems are more of an IT problem, other than what is typically considered in the literature, because in the development projects' life cycle, strong requirements should be demanded to validate the input. At the limit, they may "even find a stakeholder who is aggrieved by another stakeholder decision".

At the tool's level and after assessing various proposals and their costs, they chose Trillium, although it seems clear that this option was mainly due to costs. Our interviewee informed that MyTelecom will be looking very carefully at available open source solutions.

No formal data quality methodology [14] has been adopted, so the method they use is entirely based on intuition and "common sense".

They started the cleaning process with all clients, but then felt the need to target, "because customers do not all have the same value for the company": "fixed, identified and active customers"; "fixed, identified and inactive"; "fixed, unidentified and active (prepaid)", and so on. They have also prioritized customers' attributes, so as to give priority to the cleaning process: 1) customer code; 2) tax identification number; 3) name, address, postal code; 4) email address, etc and they are analyzing impacts, like the one of having a wrong client civil state. Costumers' addresses and postal codes are validated against Official Postal Office' reference and, when they cannot do that because the address doesn't exist in that reference (it can be a new address), and if it is related to a fiber client, they confirm it at the ground.

There is not a more or less privileged DQ dimension, as it depends on the current or future needs of each costumer. Generally speaking, the most privileged dimensions are *accuracy*, *completeness* and *relevancy*, since these data dimensions are essential for operations and decision making.

They identify DQ problems through various means, such as directly by consumers, remedy tickets, IT projects on the testing phase, and even through social networks, namely twitter. Consumers can inform the non-conformities to the application support person, that situation leading to strengthening input data validation.

Several sources of DQ problems have been identified, including data entry errors by producers, lack of data entry validations and integration of data between systems. As validations are very frowned upon by the producers, they decided to implement strong data validations, but with great usability, as eg, start giving suggestions (best matching). A centralized *rules management* has been implemented, which allows reuse of standard validation rules for the various applications, and is based on *regular expressions*. This system is being implemented in several phases and to validate various attributes, such as postal codes, dates, phone numbers, etc. The integrity rules are defined, whenever possible, at the data base management system (DBMS) level releasing, in this case, the applications from these tasks.

Because of the DQ initiative some business processes have been changed, namely the data entry validations and the data transfer between the Billing system and the Customer Relation Management[7] (CRM) system.

The organization has standards, rules and classifications for all development projects, and the standard being to use, whenever possible, classifications that cut across all the development areas. On the other hand, there are standards and classifications, general and specific, to the development of specific data models associated with the projects.

C. Motivation, Communication and Users Training

Data governance reports, which present some information quality indicators (IQ) evolution, are prepared fortnightly and sent through email on newsletter format to IT professionals and business sponsors. The IQ of an entity is currently calculated as the sum of the IQ of their attributes, with no weight. Tag clouds are used to show the attributes that are positively contributing to the quality indicator (IQ) and the ones that contribute negatively. The newsletter also provides news, such as "this issue out there": technical and management articles, and they think that this newsletter is helping a lot.

They organize *data governance awareness sessions*, where they show the results of what they are doing, currently by role type - producers, custodians and customers.

Users are trained whenever is needed. Usually a training session is organized for a new employees group or when there are new versions/significant changes in applications that will impact the way data is entered. These training sessions are important to "educate" people concerning the data entry process.

D. What Has Been Achieved So Far?

Until now 561,000 customer records have been corrected, which means 17% of all costumers. They update on the billing data base and then transfer to the other systems.

[7]One possible definition is "CRM is a technology-enabled business management tool for developing customer knowledge to nurture, maintain, and strengthen profitable relationships" [33].

A centralized *rules management* has been implemented (see above), and they are "gaining allies" among business people in various areas and hierarchical levels.

E. The Future

They are going to continue the tasks of data profiling, cleaning and enrichment, as well as the identification and modification of processes that induce data quality problems.

The big challenge is to create an environment conducive to the acceptance of a *Corporate Data Policy* and *Master Data Management.*

IV. ANALYSIS AND DISCUSSION

The main driver for MyTelecom to undertake the data quality management initiative was Costumer Data, which is in line with the most relevant areas presented in the academic [30, 31] and professional [28,32] literature. However, and because they lack a corporate data management approach, their DQM initiative is strictly related to the customer data supporting the business process of billing, and their main objectives are to reduce costs (namely post office costs), which is also in line with literature [4, 28].

By applying the framework presented in TABLE I and Fig. 1 to My Telecom's DQM initiative, we can note:

- They have began bottom-up, being only focused on a particular dataset and not having a *corporate data policy* (CDP), *strategy* (CDS) *or governance* (DG);
- Their DQM initiative is strictly focused on a specific dataset in which they are using an utility-driven approach [43];
- They are not using any formal *data quality methodology* (DQm), their DQ activities being only supported by a *data quality tool* (DQt);
- They have not implemented yet a *data steward* role (DS), but they count on a dynamic *data quality champion* (DQC).

Concerning their maturity stage, and according to [28], it seems that:

- They are mostly at the *reactive stage*, because they are fixing problems and have tools to do so, their scope is limited to a functional area, their data is siloed and they have not data stewardship;
- They are *moving to the proactive stage* because they are trying to prevent data quality problems and have a centralized rules management, standards, and classifications for all development projects;
- They are *not in the proactive stage yet*, because data are not still considered as a corporate asset and therefore they have neither corporate data definitions, nor data quality culture, data steward and data governance roles.

Given the defined objectives, the DQ initiative seems to be a success, having strong sponsorship from the CIO, which is in line with previous empirical findings. They are also gaining some sponsorship on the business side, at the directors and managers levels. They have the understanding that DQ problems are much more an "IT problem" than typically is considered in the literature. This question deserves reflection on the basis of its justification, which places the *requirements analysis teams* of the “IT Development” as a “quality negotiator” between stakeholders, although we consider, like other authors, data quality being almost a *business problem* impacting directly risk mitigation, cost control and revenue optimization [28]. Their point of view can be understood by the absence of data stewardship roles. We have caught the following IT message, concerning DQM: “We (IT) are in charge, please work with us and keep it simple”

We highlight their decision to target customers, in order to assign different priorities for cleaning and enrichment, as well as prioritizing attributes for improvement according to their usability, optimizing the utility/cost trade-offs associated with DQ improvement, which is in line with recent DQ literature [43]. Two other interesting architectural constructs are:

- The implementation of strong validations, but with very good usability, like giving suggestions;
- The creation of a centralized *rules management* repository that, in addition to helping the data quality improvement through centralized validations, facilitates greater productivity of development teams.

Their biggest focus is on motivating employees through innovative forms of communication, in addition to training, which appears to be a critical success factor [5] (CSF) for data quality management. The characterization of the quality indicators, and the dissemination of its biweekly trends in newsletter format using tag clouds, has had excellent results with regard to motivation and involvement of multiple users. Another positive aspect is the recent introduction, for IT people, of a “data quality” key performance indicator[8] (KPI), despite not having a large weight.

Another CSF seems to be the data quality management leadership by a person with a data quality champion profile [3].

Even though this case study cannot be generalized for populations or universes [10], it shows the existent gap between DQ research efforts that develop and enforce the application of DQ methodologies for quality assessment and improvement, and what is actually done in the practice.

The available literature, particularly from the professional side, advises that a data quality management environment should be designed top-down, traversing strategic, tactic and operational levels in that order. It appears that the way MyTelecom is working is clearly bottom-up, although counting on the CIO commitment, thus leaving unanswered the question:

Can an organization achieve a high maturity level in its DQM initiative just beginning it bottom-up, by working a cost-effective business case, and thus lead the organizational awareness about the quality of its data, that inducing a top-down strategy to be followed by some iterations either top-down or bottom-up? Is that approach common in the business world?

[8]According to [35] KPIs are compilations of data measures used to assess the performance of business´s operations.

V. Limitations and Future Work

Our research questions being *why* a specific company decided to embark in a data quality management initiative, *how* it did that and *which objectives* have been achieved, so far, as well as comparing what we found in the case to theories from the literature have been achieved and, from the case study’s findings, others have appeared. This work being a single explanatory case study, it only aimed to describe a real situation in its context and validate the findings against existing theory. In fact, to our best knowledge [3, 34], very few case studies concerning this broad issue are available, so we tried a small contribution to this knowledge.

This effort will continue through multiple case studies, crossing different businesses, possibly supplemented by a survey addressed to business and IT people, in each organization, if we can find available resources and organizations willing to welcome these studies.

In any case, it is decided to maintain and deeply analyze the research questions underlying this work, switching to one of the research based Data Quality Management Maturity Models, because they are rigorously defined and have an underlying theory base.

VI. Acknowledgments

We would like to thank our anonymous referees for their helpful comments on the first version of this paper.

References

[1] I. Benbasat and R. W. Zmud,“Empirical Research in Information Systems”, MIS Quaterly, vol. 23, no. 1, pp. 3-16 , 1999.

[2] M. Rosemann and I. Vessey, I., “Toward Improving the Relevance of Information Systems Research to Practitioners: The Role of Applicability Checks”, MIS Quaterly, vol. 32, no.1, pp. 1-22 , 2008.

[3] S. W. Tee, P. L. Bowen, P. Doyle and F. H. Rohde, “Factors influencing organizations to improve data quality in their information systems”, Accounting and Finance, vol. 47, pp. 335-355, 2007.

[4] L. English, Improving Data Warehouse and Business Information Quality. Wiley & Sons, 1999.

[5] J. K. Leidecker and A.V Bruno, “Identifying and using critical success factors, Long Range Planning, vol. 17, no. 1, pp. 23-32, February 1984.

[6] The Data Warehousing Institute, Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data. Seattle, WA: TDWI, 2002.

[7] J K. Weber, B. Otto, and H. ¨Osterle, “One size does not fit all—a contingency approach to data governance”, ACM J. Data Inform. Quality, vol. 1, no. 1, June 2009.

[8] A. Bitterer, and D. Newman, Organizing for data quality. Stamford, CT: Gartner Research, 2007.

[9] T. Friedman, Gartner study on data quality shows that IT still bears the burden, Stamford, CT: Gartner Group, 2006.

[10] R. K. Yin, Case Study Research: design and methods, 3rd ed. Thousand Oaks, CA: Sage Publications Inc, 2003.

[11] IBM Data Governance, viewed 23.01.2009, at http://www-01.ibm.com/software/tivoli/governance/servicemanagement/data-governance.html

[12] HP, Managing data as a corporate asset: three action steps toward successful data governance, White Paper, 2007, viewed 11.05.2009, at http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA1-2494ENA.pdf

[13] DAMA International, The DAMA Guide to Data Management Body of Knowledge. Bradley Beach, NJ: Technics Publications LLC, 2009.

[14] Batini, C. Cappiello, C. Francalanci and A. Maurino, "Methodologies for data quality assessment and improvement", ACM Comput. Surv, vol. 41, no.3, July 2009.

[15] C.Batini and M. Scannapieco, Data Quality: Concepts, Methodologies and Techniques, Springer-Verlag, 2006.

[16] T. Friedman and A. Bitterer, "Gartner Magic Quadrant for Data Quality Tools", 23 June 2009, viewed 29.11.2009 at http://www.gartner.com/technology/media-products/reprints/dataflux/167657.html

[17] J. M. Juran, Juran on Leadership for Quality: An Executive Handbook. New York: The Free Press, 1989.

[18] J. M. Juran and F. M. Gryna, Quality Planning and Analysis, 2nd ed. New York: McGraw-Hill, 1980.

[19] T. C. Redman, Data Quality The Field Guide. The Digital Press, 2001.

[20] Y. Wand and R. Wang,"Anchoring Data Quality Dimensions in Ontological Foundations", Communications of the ACM, pp. 86-95. November 1996.

[21] R.Y. Wang and D.M.Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers", Journal of Management Information Systems, vol. 12 , no. 4, pp. 5-34, 1996.

[22] Y. Wand and R. Y. Wang, "Anchoring Data Quality Dimensions in Ontological Foundations", Communications of the ACM, vol. 39, no. 11, pp. 86-95,1996.

[23] T. C. Redman, Data Quality for the Information Age. Artech House, 1996.

[24] I. Caballero, A. Caro, C. Calero and M. Piattini, "IQM3: information quality management maturity model", Journal of Universal Computer Science, vol. 14, no. 22, pp. 3658-3685, 2008.

[25] K.M. H¨uner, M. Ofner and B. Otto, "Towards a maturity model for corporate data quality management" in Proceedings of the 2009 ACM symposium on Applied Computing (Honolulu, Hawaii), New York: ACM, 2009, pp. 231-238.

[26] K.-S. Ryu, J.-S. Park and J.-H.Park, "A data quality management maturity model", ETRI Journal, vol. 28, no.2, pp. 191-204, 2006.

[27] IBM. The IBM data governance council maturity model: Building a roadmap for effective data governance, Technical report. IBM Software Group, 2007.

[28] T. Fisher, The Data Asset – How Smart Companies Govern their Data for Business Success. Hoboken, NJ: John Wiley & Sons, Inc, 2009.

[29] D. M. Strong, Y. W. Lee and R. Y. Wang, "Data Quality in Context", Communications of the ACM, vol. 40, no. 5, pp. 103-110, May 1997.

[30] B. Otto, K. Wende, A. Schmidt and P. Osl,"Towards a framework for corporate data quality management" in Proceedings of the 18th Australasian Conference on Information Systems, pp. 916-926.

[31] A. Umar, G. Karabatis,L. Ness, B. Horowitz and A. Elmagardmid, "Enterprise Data Quality: A Pragmatic Approach", Information Systems Frontiers, pp. 279-301, 1999.

[32] Informatica, Addressing Data Quality at the Enterprise Level,White paper, 2005, viewed 23.02.2009, at http://www.i2.com/partners/Partner_collateral/INFA_DQ_wp_6631_lo.pdf

[33] P. Sue and P. Morin, A Strategic Framework for CRM. CMC–Insight Exec, Bristol, UK, 2001, viewed 21.01.2009 at www.insightexec.com/cgibin/library.cgi?action=detail&id=1385.

[34] P. Missier, G. Lalk, V. Verykios, F. Grillo, T. Lorusso, and P. Angeletti, "Improving Data Quality in Practice: A Case Study in the Italian Public Administration", Distrib. Parallel Databases, vol. 13, no. 2 ,Mar. 2003.

[35] V. Veerawat, A. Koronios and J. Gao, "A Framework for the Development of the Business Case for the Introduction of Data Quality Program Linked to Corporate KPIs & Governance, in Proceedings of the Fourth INTERNATIONAL Conference on Cooperation and Promotion of Information Resources in Science and Technology, IEEE, 2009, pp. 230-235.

[36] V. Khatri and C. V. Brown, "Designing Data Governance", Communications of the ACM, vol. 53, no. 1, pp. 148-152, January 2010.

[37] J.A.Hoffer, M.B. Prescott and F.R. McFadden, Modern Database Management, 8^{th} edition. Upper Saddle River, NJ: Pearson Education Inc, 2007.

[38] P.F. Drucker, The New Realities. New Brunswick, NJ: Transaction Publishers, 2003.

[39] International Standards Organization. ISO 9000:2005 (E) - Quality Management Systems: Fundamentals and Vocabulary, 3rd edition, ISO, 2005.

[40] Dama International.The DAMA Dictionary of Data Management. Technics Publications LLC, NJ, 2008.

[41] A. Lucas, A. Palma-dos-Reis and M. Caldeira. "The Quality of Monitoring Data in Civil Engineering Works",in Proceedings of the 14th International Conference on Information Quality (IQ 2009), 2009.

[42] L.English, Improving Data Warehouse and Business Information Quality. Wiley & Sons, 1999.

[43] A. Even and G. Shankaranarayanan, "Dual Assessment of Data Quality in Customer Databases, ACM J. Data Inform. Quality, vol. 1, no. 3, December 2009.